Addressing NYC Transit Deserts Through Local Self-Organized Commuter Vans

Final Project Report

for the Urban Science Intensive (USI II), Summer 2019

Evgeniya Bektasheva, Hanxing Li, Kenji Uchimoto, Keundeok Park, Wenjie Zheng

Center for Urban Science and Progress

New York University

NY, USA

Abstract

While New York City has developed public transportation networks provided by MTA, a huge number of people living in outer boroughs are left outside the networks. These people are living in “transit deserts” and have a strong demand for transportation means. Part of the demand is met by so-called “dollar vans” – a chain of thousands of privately owned commuter vans which run across the NYC. A startup called “Dollaride” connects drivers and passengers in these marginalized communities using an innovative transportation technology.

The project identifies methodology for calculating transit deserts (transit-underserved areas in terms of specific transit supply-demand equilibrium) in a large city. The methodology is based on earlier developed attitude by Jiao and Dillivan (2013), but some different parameters characterising a first-tier American city were suggested. The outcome of the project is a map of transit deserts in NYC, based on a solid quantitative analysis and data. Another practical implication of the project is a list of new routes for Dollaride, which would benefit its passengers, drivers and NYC transportation authorities.

Keywords: Transit Desert, Commuter Van, Transportation, Urban, Dollar Van, Routes

Introduction

The Metropolitan Transportation Authority, commonly known as MTA, runs the largest mass transit in the United States. Among the top of the world in terms of annual ridership, system track mileage, and a number of stations, the system is also known for running 24/7. However, a large percentage of the population still lacks a convenient and viable means of transportation for their daily commute. The gaps in the MTA service network and the need for a proper transit gave rise to a system of private commuter vans, also known as “dollar vans”.

Dollar vans have played a very important role in NYC's transportation system for over 30 years. They usually have a seating capacity of around 20 passengers. The vans provide alternative transit options with frequent services that traverse several major corridors and neighborhoods in the outer boroughs of the city. Although the commuter van system has been in place since the 1980s, it is still very much unregulated and undeveloped.

The transportation system in New York needs to improve to become faster, more cost-effective, and being able to serve the people who depend on it the most. Consequently, dollar vans appear as an important supplement of mass transit that helps to fill the gap between transit deserts and transit-rich areas. Dollaride, as a new player on the transportation area, can add value by connecting passengers and dollar vans in transit deserts using technologically advanced attitude. Moreover, by suggesting new routes using a data-driven way - rather than random historical trips used by local residents - Dollaride can improve the transportation situation in NYC.

Historically, the public transit system in the US has been oriented towards private vehicles. The government initially failed to create a balanced system that would serve a large amount of transit-dependent population in low-income, central-city areas (Garrett and Taylor, 1999; Campanella, 2018). Therefore, one of the major objectives of the transit system researchers was finding a way to improve that, as well as identify “transit deserts” within particular city. After detailed analysis of the literature (see Appendix 2, Literature Review), it became clear that existing methodology does not take into account a substantial increase in population density, congestion, dependence on certain transportation means (such as subway) and common “last mile” commutes from a subway to a final destination in large cities (Chavis and Daganzo, 2013). Therefore, an amended methodology should be created in order to identify transit deserts and optimal routes in New York City.

Problem Statement

1. ***Define "transit desert" and identify the location of the deserts in NYC.***

While Jiao and Dillivan (2013) defined transit desert as "areas that lack adequate public transit service", we will elaborate on this definition and provide a quantitative assessment of those areas. The final definition of the transit desert will be assessed for the applicability of the New York City specifically. Final outcomes would include a contour/heat map of NYC, visualizing the different levels of severity and demands for public transportation by census-based areas. Such map can be used for the public policy needs and outline the areas of improving the transportation network in NYC.

2. ***Propose new routes for Dollaride in areas with high demand for public commute.***

Taxi, FHV (For Hire Vehicles) and private cars can instantly serve people and become a practical solution for those located in transit deserts. Consequently, they could be viewed as indirect competitors or substitutes to private vans - Dollaride clients. Therefore, analysis of taxi and FHV data (pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, rate types, and driver-reported passenger counts) can help to create a map of the most common trip records. This map would also use population data, counting the number of people living and travelling between zones.

As a next step, such analysis could be a good identifier for new routes for private dollar vans. To achieve this, ranking the routes based on the total number of trips would be conducted using statistical methodology. Proposed new routes will be identified for Queens and Brooklyn, which are two major markets for Dollaride.

Data

There are three categories of data that were used for this project: publicly available datasets, proprietary information, and data collected as part of the project discovery. Most of the research is done using publicly available data sources. Thus, we used census tract data including demographic information by area; locations of MTA subway/bus stations and schedules trains/buses from NYC Open

Data and MTA GTFS; taxi and commuter van data from NYC DOT & TLC and other data sources (see Appendix 1, Data Sources).

Then, the synthetic population trip data was provided by NYU C2SMART. The dataset is generated from a survey that asks people about their transit needs and their transit routines. It contains estimated personal trip data for 8+ million population in New York City on a weekday, and every record has information about trip origination, trip destination, trip start time, trip end time, purpose of the trip, and mode of transportation, etc.

Finally, collecting a first-hand commuter van trip data was also necessary for this project due to the lack of existing data and the necessity of understanding commute van operations. Supply of dollar van services was assessed through two methods: the dollar van customer and drivers' survey and collecting GPS ridership data during several field trips organised by the team (Appendix 3, Field Trip and Dollar Van Supply Assessment). A rudimentary assessment of serving the potential neighborhood, with demand for commuter vans, was conducted with the aforementioned data.

All of the data used for the project have certain limitations. Publicly available datasets have aggregate data and therefore certain information can be missed. Proprietary data provided by NYU C2SMART was limited to certain variables; it was also computationally heavy and required a lot of time and resources to process. Finally, field trip data was limited to the number of trips, resources and time constraints of the project.

Methodology

1.Methodology for identifying "transit deserts" in NYC

When defining the *transit deserts* in New York City in terms of *transportation supply* estimation, we are looking at the MTA service coverage and performance, using the NYC Open Data and GTFS (Google's General Transit Feed Specifications) datasets. These datasets contain geographical information about an MTA bus line or subway line stops and routing. Based on this information, we are able to visualize the coverage of MTA services. We will consider the closest subway station or MTA bus stop, and identify the distance from them to the centroid of the census tract (being a theoretical distance an average person is willing to walk to the transit point). In terms of *transportation demand* estimation, population data was used (with a granularity of a census tract, as the most computationally small data available). After that, for the purpose of creating an index of each transportational zone, z-score of the distance variable (supply) and z-score of the population density variable (demand) was calculated. The methodology assigned 50% to the combined MTA distance factor (taking the negative value of the score, meaning the further the subway/bus stop is from the area, the more requirement for transportation is for that area) and 50% weight to the population variable (meaning that those are the areas in a high traffic zone). The logic behind the chosen parameters and the weights for them, while considering transit deserts in a large megapolis like NYC, is the following. Various research (Jiao and Dillivan, 2013; Jiao, 2018) used the same coefficients for each variable they considered when defining transit deserts, but they worked with data from low density cities, which had no major public transportation means - such as the subway in NYC. However, urban research (Garrett and Taylor, 1999) suggested that the public transit system in any large city is an important factor and even a source of social inequality, because resources should be allocated to improving public transit service in low-income, central-city areas serving a high proportion of transit dependents. Therefore, the proposed methodology is giving more weight to the subway stop distance parameter and the population density variable:

$$\textit{Index} = 0.5 \times \textit{Population\_density} + 0.4 \times (-1)\ \textit{Distance from a Subway Stop} + 0.1 \times (-1)\ \textit{Distance from a Bus Stop}$$

Index for each location was calculated in the range from “1” to “4”, with “4” being the area with the highest requirement for the public transportation (a transit desert). The analysis of most needed areas in NYC was performed and mapped.

2.Methodology for identifying new optimal routes for Dollaride

While providing new Dollaride routes for people living in transit deserts is important, those routes should be sustainable from a business perspective. Dollar vans are operated by private companies and have to be profitable. By applying the Vehicle Routing Problem, the optimal set of routes for a fleet of vehicles to traverse in order to deliver to a given set of customers (Chakroborty, P., Deb, K. & Subrahmanyam, P., 1995), we will create new routes which maximize number of potential passengers for the routes.

First step for this task is to analyse the taxi data and FHV (For Hired Vehicle) data, released by NYC OpenData portal. Using FHV trip data, trips with the number of passengers greater than or equal to two were selected. The logic here is that customers of FHV shared ride services (such as Uber Pool or Lyft shared) are potential customers of Dollaride as they share the ride and use app-based service. Manhattan was excluded, since Manhattan is not in Dollaride coverage and therefore not relevant for the analysis.

Then, the numbers of trips is aggregated based on combinations of origin and destination zone. Using the aggregated trip data, we combine consecutive 9 taxi-zones. The combined zones become “routes”, and the goal is to find the best routes in which dollar vans can pick up the highest number of passengers. For that, first, we pick an origin zone. Then, from the adjacent zones, we choose a first destination zone along with which we can maximize the number of passengers. In the Figure 1, Zone A is the origin and the first destination should be Zone D. The number in the arrow means the number of passengers who move from one zone to another. The number inside of a zone symbol is the number of passengers who travel within the zone (intra-zone trip). Then, starting from the Zone D, the algorithm chooses Zone F over G and H. The algorithm repeats this process 8 times each.

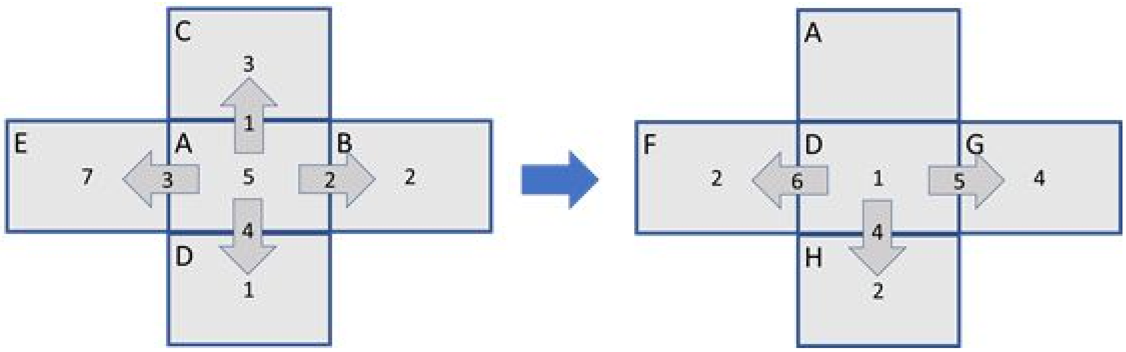


Figure 1: Graphical description of the best route methodology

The limitations of this method is that we only use nine zones, and this number was chosen subjectively due to time constraints and computational resources. The method could be improved further by using larger data.

Results

1. *Analysis of Transit Deserts*

The analysis of the transit deserts in NYC shows that the neighborhoods with limited transit accessibility include Harlem in Manhattan, Long Island City in Queens and Bushwick, parts of Flatbush, Sunset Park and most of Coney Island in Brooklyn, as well as East Brooklyn. From a public policy perspective, the city should develop a roadmap to connect those communities: either by new subway construction or by other means of public transportation, including more extensive bus service. Harvard researchers revealed that good transportation is a critical factor for upward economic mobility (Chetty and Hendren, 2015).

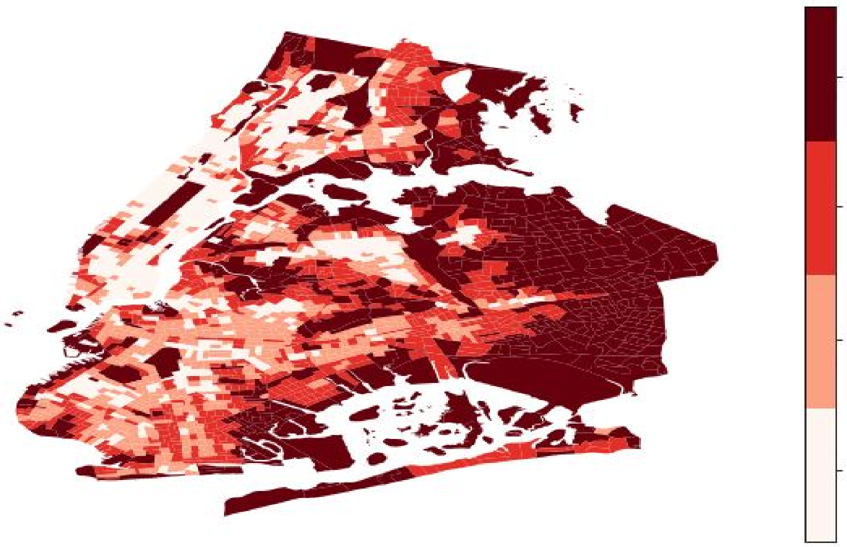


Figure 2: Transit deserts in NYC

After the transit desert map was created, we also assessed whether the transit-underserved areas overlap with the areas with (a) larger amount of people driving to work every day and (b) limited bus supply (the analysis was based on the MTA real-time data) (see Appendix 4, Supplementary Data Analysis). Indeed, the transit desert areas were similar to the areas with larger percentage of commuters and less frequent bus service (Figure 3). Some portions of the city are not connected to subway service and are left to rely on cars or long walks to get to subways or take unreliable bus service. Some parts of Queens and Northern/East Brooklyn with a large percentage of commuters (Figure 3, left) also had infrequent bus service; however, some central areas in Queens had more frequent bus service. Analysis of the Google map revealed that it was the areas with major job centers, such as Flushing, East Williamsburg/Maspeth IBZ, etc. Such supplementary analysis supported our methodology and findings of the transit deserts in NYC.

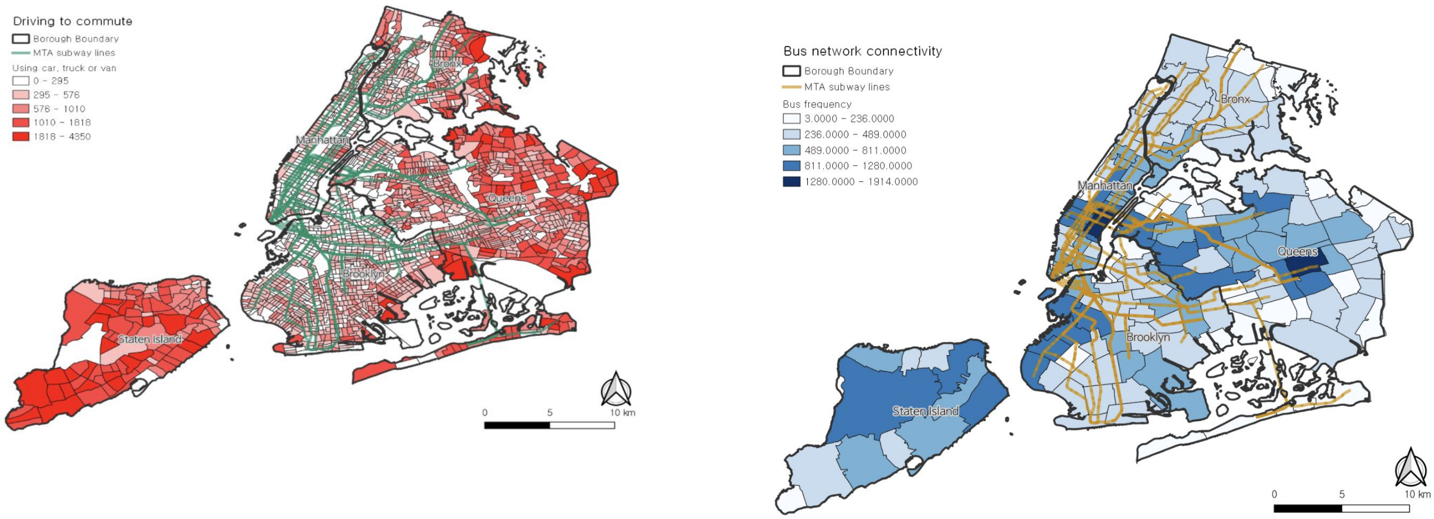


Figure 3: Heatmaps of a number of people driving to commute (left) and bus frequency(right)

2. FHV Data analysis and Optimal routes for Dollaride

As the first step of the taxi and FHV data analysis, we looked at the different trip pattern between weekdays and weekends for trips taken by 100 people and more, statistically significant number of people identified in the course of research (Figure 4). Analysis of Figure 4 shows that a large amount of the FHV trips occurred at the area nearby Williamsburg, East Brooklyn during weekends. Interestingly enough, the pattern was observed in the areas some of which were identified as transit deserts (Figure 2).

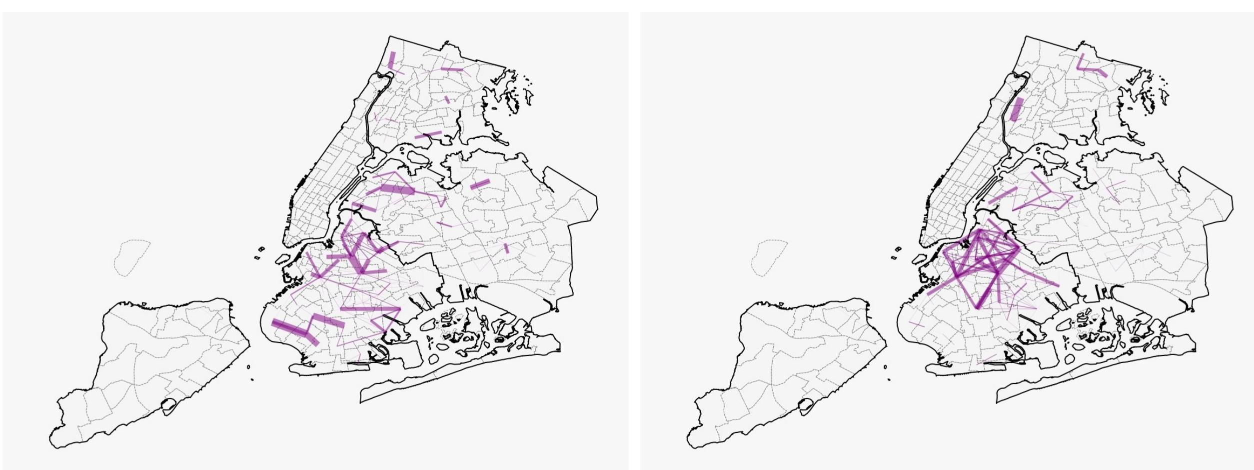


Figure 4: Weekday FHV trip(left) and Weekend FHV trip(right) in 2018 06:00 am to 10:00 pm

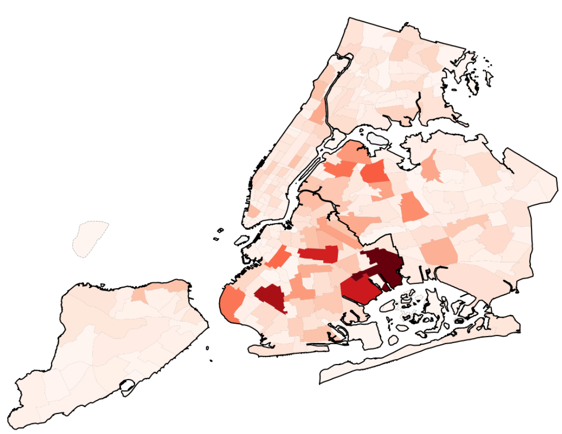


Figure 5: Heatmap of the number of internal FHV trip

Additionally, the largest trips are generated inside of the taxi zones. Since we cannot identify the location in the same taxi zone, origin and destination plotting map does not represent them so that we have the heatmap for the inside trip of FHV(Figure 5). Maps represent no different pattern between weekday and weekend trip. The area that has the largest amount of the internal trip is East NewYork which is located in the eastern part of Brooklyn. The Google analysis revealed that the gateway center shopping mall is located in this area (which does not have any public transportation connections).

After that, when the above two datasets were merged and the numbers of trips are aggregated based on combinations of origin and destination zone and proposed statistical methodology, we obtained a map of top 20 proposed new optimal routes for Dollaride. For this purpose, only Queens and Brooklyn was considered (being the main markets for the dollar van drivers in NYC). These new routes were suggested because of the sufficient demand and lack of public transportation supply. For the top 20 optimal routes, 11 of them are located in Queens and 9 in Brooklyn. Figure 6 and 7 show the areas where the optimal routes could be made, with the yellow line showing NYC subway lines. Again, the majority of them overlap with the transit deserts identified earlier in the research (Figure 2).

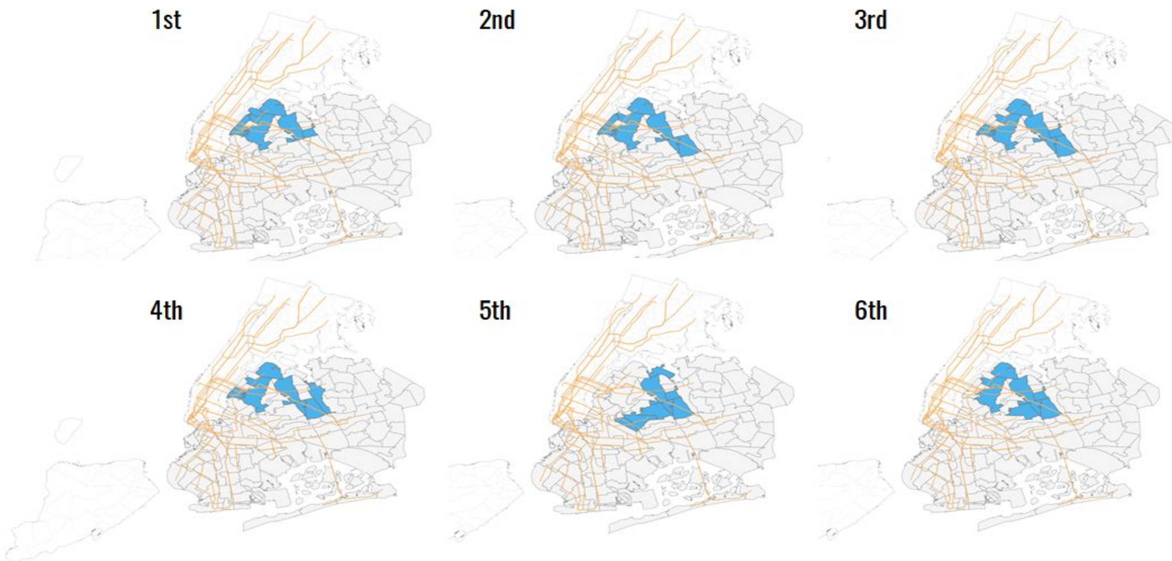


Figure 6: Suggested areas for the optimal Dollaride routes in Queens

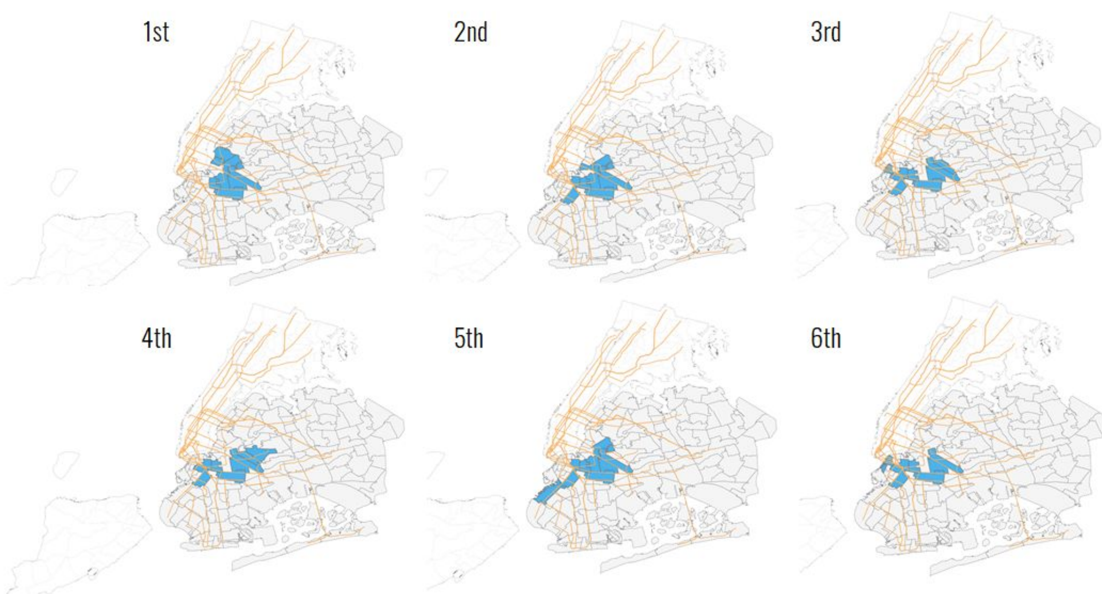


Figure 7: Suggested areas for the optimal Dollaride routes in Brooklyn

Finally, analysis of the synthetic population data provided by NYU C2SMART also revealed an interesting pattern. Since it describes the general transit behavior of the population, we looked at the taxi zones with highest number of trips as origin and destination, also only in Brooklyn and Queens (Figure 8). Due to the large amount of dataset, only ten percent of data was used. Again, some of the neighborhoods which are marked dark blue (parts of Flatbush, Forest Hills, Borough Park) were also identified as transit deserts in Figure 2. Some further analysis is described in Appendix 5 (Synthetic Population Data Analysis for Specific Areas in NYC). The analysis indirectly supports our choice of transit deserts in NYC.

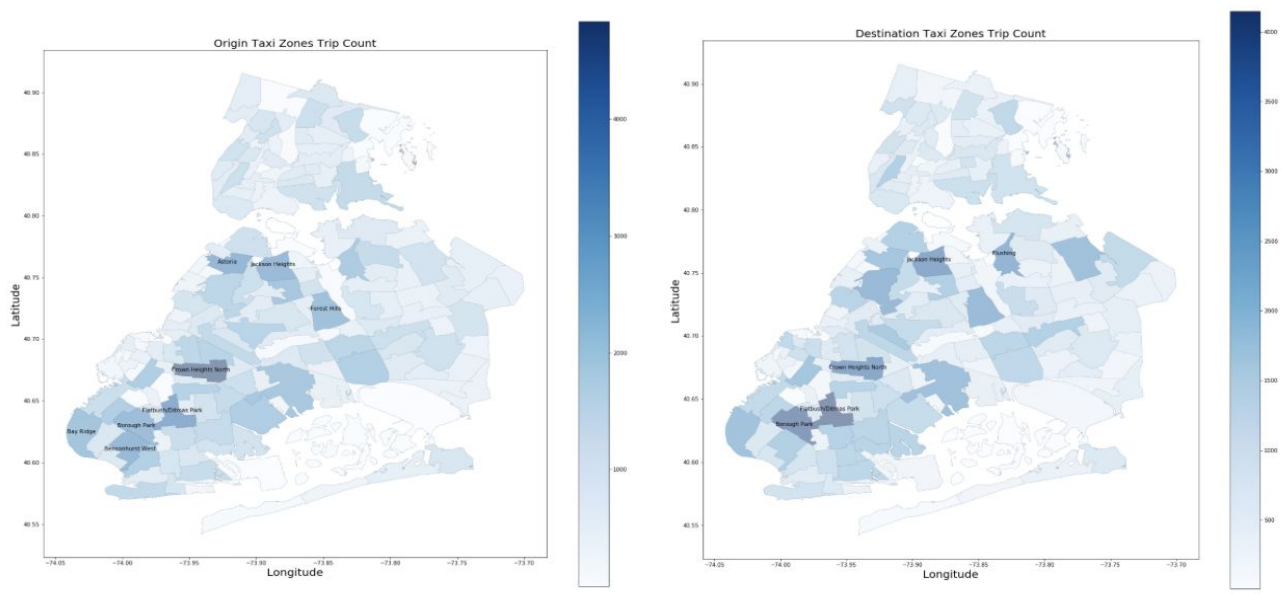


Figure 8: Heatmap of Origin and Destination for Synthetic Population Trips in NYC

Conclusions

The research focuses on both public policy and private business aspects of resolving the transportation problem in NYC. Suggested methodology on identifying transit deserts was tested using other datasets. It could be improved by using more variables and adding more granularity. On the other hand, the city has limited resources and budget to reshape and improve transportation services, and private vans can add a lot of value by serving the needs of people in NYC.

References

[1] Jiao, J., & Dillivan, M. (2013). Transit Deserts: The Gap between Demand and Supply. *Journal of Public Transportation*, *16*(3).

[2] Chang, S. K., & Schonfeld, P. M. (1991). Multiple Period Optimization of Bus Transit Systems. *Transpn*, 453-478.

[3] Chavis, C., & F.Daganzo, C. (2013). Analyzing the structure of informal transit: The evening commute problem. *Research in Transportation Economics*, 277-284.

[4] Goldwyn, E. (2018). Anatomy of a new dollar van route: Informal transport and planning in New York City. *Journal of Transport Geography*.

[5] King, D. A., & Goldwyn, E. (2014). Why do regulated jitney services often fail? Evidence from the New York City group ride vehicle project. *Transport Policy*, 186-192.

[6] Kittelson & Associates, Inc.; Parsons Brinckerhoff, KFH Group, Inc.; Texas A&M Transportation Institute; Arup; Transit Cooperative Research Program; Transportation Research Board; National Academies of Sciences, Engineering, and Medicine. (2013). Ch 5: Quality of Service Methods. In *Transit Capacity and Quality of Service Manual, Third Edition* (pp. 6-130). Washington D.C.: The National Academies Press.

[7] Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., & White, P. (2004). *The demand for public transport: a practical guide.* London: TRL Limited.

MTA; NuStats. (2009). *2008 New York Customer Travel Survey Final Report.* Austin, Texas: NuStats.

[8] Tirachini, A. (2013). The economics and engineering of bus stops: Spacing, design and congestion. *Transportation Research Part A*, 37-57.

[9] Campanella, J. (2018). Watering New York's Transit Deserts. *The New York Times*. Retrieved April 27, 2019 from https://www.nytimes.com/2018/08/13/opinion/mta-new-york-transit-deserts.html.

[10] Pratt Center for Community Development. (2012). Transit Access for the Greatest Need New Yorkers. Retrieved April 27, 2019 from https://prattcenter.net/transit-access-greatest-need-new-yorkers.

[11] Garrett, M., & B. Taylor. (1999). Reconsidering social equity in public transit. *Berke ley Planning Journal* 13: 6-27.

[12] Train C. (2018). The Economic Toll of New York's Inaccessible Subway. Retrieved April 28, 2019 from https://www.citylab.com/transportation/2018/08/half-of-nyc-neighborhoods-are-transit-deserts/566471/

[13] Stringer, S. (2018). Service Denied: Accessibility and the New York City Subway System. NYC Comptroller report. Retrieved April 28, 2019 from https://comptroller.nyc.gov/reports/service-denied-accessibility-and-the-new-york-city-subway-system/

[14] Access Denied. Report by the Transit Center (2018). Retrieved April 28, 2019 from https://transitcenter.org/publications/access-denied/#introduction

[15] Jiao, Junfeng & Dillivan, Maxwell. 2013. Transit Deserts: The Gap between Demand and Supply. Journal of Public Transportation, 16 (3): 23-39.

[16] Jaffe, E. (2015). Where the New York City Subway Doesn't Go. Retrieved April 28, 2019 from https://www.citylab.com/transportation/2015/08/where-the-new-york-city-subway-doesnt-go/400538/

[17] Chakroborty, P., Deb, K. & Subrahmanyam, P. (1995). Optimal Scheduling of Urban Transit Systems Using Genetic Algorithms. *Journal of Transportation Engineering, Vol. 121, Issue 6 (November 1995).* Retrieved July 20, 2019 from http://faculty.salisbury.edu/~ajlembo/419/networkpaper.pdf

[18] Chetty and Hendren, 2015. The Impacts of Neighborhoods on Intergenerational Mobility. Retrieved July 20, 2019 from http://www.equality-of-opportunity.org/images/nbhds_exec_summary.pdf

Appendix 1

Data Sources

[1] Traffic: New York Department of Transportation. https://www.dot.ny.gov/tdv

[2] FHV data https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

[3] Zip Code - New York City.

https://data.cityofnewyork.us/Business/Zip-Code-Boundaries/i8iw-xf4u/data

[4] Accessible Stations in the MTA Network, MTA http://web.mta.info/accessibility/stations.htm; MTA Static GTFS Data Feeds, http://web.mta.info/developers/

[5] Customer and driver survey & interview from Dollaride

[6] MTA subway/bus stop locations and frequency http://web.mta.info/nyct/subway/howto_sub.htm

[6] Census Tract Data: the New York City Department of City Planning.

https://data.cityofnewyork.us/City-Government/2010-Census-Tracts/fxpq-c8ku

[7] Commuter van driver data from NYC DOT & TLC

[8] Synthetic population trip data from NYU C2SMART

[9] MTA bus GTFS: Open Mobility Data. https://transitfeeds.com/p/mta

[10] Route Report Card : Turnaround. http://busturnaround.nyc/report-cards/

Appendix 2

Literature Review

Historically, the public transit system in the US has been oriented towards private vehicles. As Garrett and Taylor (1999) noted, the public transit system, being an important part of any large city, were mostly oriented to two principal markets - "downtown commuters and transit dependents - people who are too young, too old, too poor, or physically unable to drive". The paper argues that social inequality in public transit exists because comparatively few resources were allocated to improving public transit service in low-income, central-city areas serving a high proportion of transit dependents. Correspondingly, transportation planners should re-examine current public transit policies and plans.

This statement becomes even more valid due to a substantial increase in population density, congestion and parking fees in metropolitan cities. Arguably, the two markets identified by Garrett and Taylor merged nowadays. According to the research by Pratt Center for Community Development in 2012, nearly two-thirds of 750,000 New Yorkers whose commute to work takes over an hour, have an annual family income under $35,000. And a lot of areas in Brooklyn and Queens lack reliable transit and unserved by MTA. For example, some neighborhoods in southeast Brooklyn are located more than three miles from stations, and there is only a single subway line bringing residents to the center. Plans for extending subway service to this part of Brooklyn were drafted back in the 1920s but had never been executed. Thus, in his recent research, Campanella (2018) noted that "today, with urban living resurgent and transit-rich parts of New York City becoming increasingly unaffordable, we need to revisit plans to better link boroughs outside Manhattan to the city's center".

Another aspect is transportation accessibility for the mobility-impaired people. Only 24 percent of New York City's 472 subway stations are accessible via an elevator, according to a 2018 report by the City Comptroller's Office, and half of the city's subway-served neighborhoods qualify as "ADA transit deserts," meaning that they lack a single accessible station. As Train (2018) stated, "In these areas, nearly 640,000 residents are impacted, including those who are mobility impaired, seniors, and children under five, who often need a stroller”.

Therefore, one of the major objectives of the transit system researches is finding a clear definition of the term “transit desert” and providing geographical identification of those areas. Jiao and Dillivan (2013) implemented a quantitative approach in defining transit deserts. “The term “transit desert” is a new concept that looks at the gap between the level of transit service (supply) and needs of a particular population (demand). These populations are often referred to as “transit dependent,” people that are too young, too old, or too poor or who are physically unable to drive. “Transit deserts” in this case are defined as areas that lack adequate public transit service given areas containing populations that are deemed transit-dependent”. Thus, the criteria and formula are then designed to calculate the supply and demand and transform them to z-scores with a set of data sources including the demographic data from 2010 U.S. census and transit frequencies and transit stops distribution from the GTFS (Google's General Transit Feed Specifications). The transit supply is defined by 4 criteria, which are the number of bus and rail stops in each census block, the frequency of service for each bus and rail stop per day in each census block, number of transit routes in each census block, and length of bike routes and sidewalks in each block since biking and walkability is also assumed to be aiding the transit availability by the Jiao and Dillivan group. The data for each of these variables was standardized and an overall z-score was calculated. A positive and higher value of z-score suggested that a certain census tract is well-served by the public transit while a lower and negative value of z-score would suggest that the public transit needs to be improved.

Appendix 3

Field Trip and Dollar Van Supply Assessment

The quantitative analysis isn't the sole approach for this project. It should be accompanied by local knowledge and experiences from drivers and passengers, as indicated in commuter van research (Goldwyn, 2018). These qualitative and quantitative approaches will together strengthen the decision making of better serving the communities.

Therefore, several field trips were organized in order to assess the supply of dollar van market in terms of the number of routes, drivers and vans, scheduling, the number of vehicle trips, and service coverages. It was important to understand that some routes may have not enough vans and other - excessive supply with vans often run empty during certain hours, meaning overcrowding and/or underperformance. Such qualitative analysis was essential to trying to arrive at the equilibrium of the dollar van market.

The team used a publicly available app and gathered information about the eight existing dollar van routes (Figure 1). In Brooklyn, there are two dollar van routes. One operates along Flatbush Avenue between downtown Brooklyn and the King's Plaza Shopping Center, and another one operates on Utica Avenue between Eastern Parkway and the King's Plaza Shopping Center. There are three routes in eastern Queens operating out of the Jamaica Station and connect various neighborhoods, communities, and the Green Acres Mall in easteen Queens. The passengers can often get on and get off the van at any location along the route. In comparison to the aforementioned routes that mainly serve the purpose of connecting the transit hub and communities, the 3 dollar routes which connect three Chinatown in Manhattan, Brooklyn, and Queens operate a different model. They make direct and nonstop service between terminal and terminal including Sunset Park in Brooklyn, Chinatown in Manhattan, and

Flushing in Queens. In speaking of the Flatbush Avenue route, though the commuter vans do not follow a fixed schedule, they come at a pretty high frequency and multiple vans were observed passing by within a short period of time at around 5 pm during the rush hour. The terminal-to-terminal travel time for the dollar van is around 1 hour for the 7-mile long route between Smith Street and King's Plaza Shopping Center. Overall, the dollar van offers a pretty convenient transit option compared to the MTA B41 bus whose route overlaps with the commuter van route most of the time. The reason is probably because MTA B41 has a poor on-time performance of 46%, and travels at a slightly lower average speed at 5.8 miles per hour [11]. The dollar van is also 27% cheaper than the MTA fare. From the curbside interview with several passengers, a guaranteed seat, cheaper fare, and shorter duration compared to the MTA bus were among the top reasons they choose to take the commuter van.

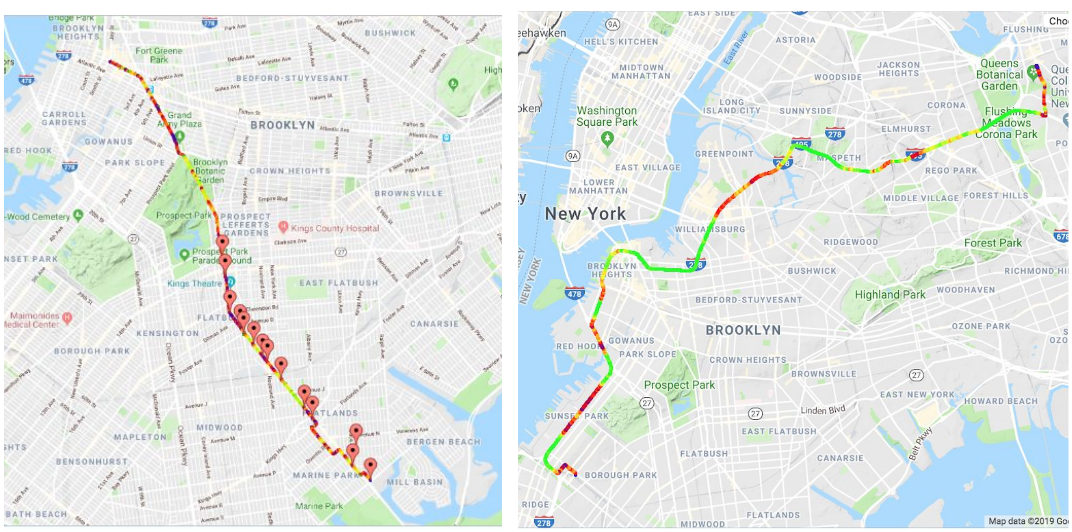


Figure 1: Sample Dollar Van Routes from GPX Data

Appendix 4

Supplementary Data Analysis

A. Private commute data analysis

According to MTA data, there are currently 468 subway stations in NYC that serve around 5.6 million daily riders**.** Yet, there is a large percentage of commuters drive to work every day, and most of them live in outer boroughs in Staten Island, east Queens, south Brooklyn and north Bronx, which are represented by the dark red areas on the map (Figure 1)

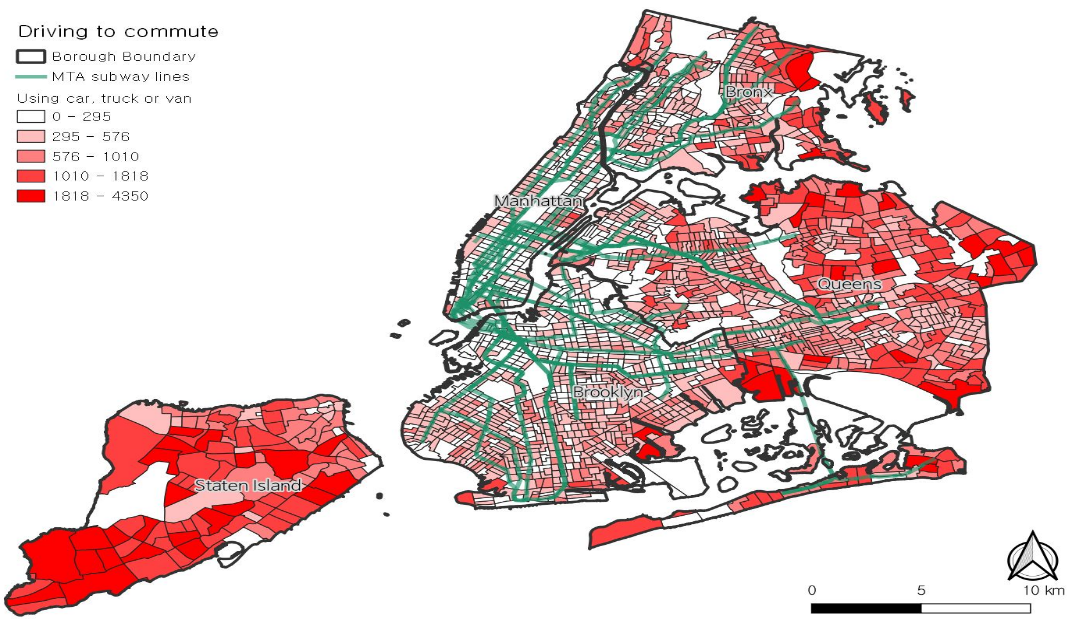


Figure 1: Heatmap of number of people driving to commute in NYC

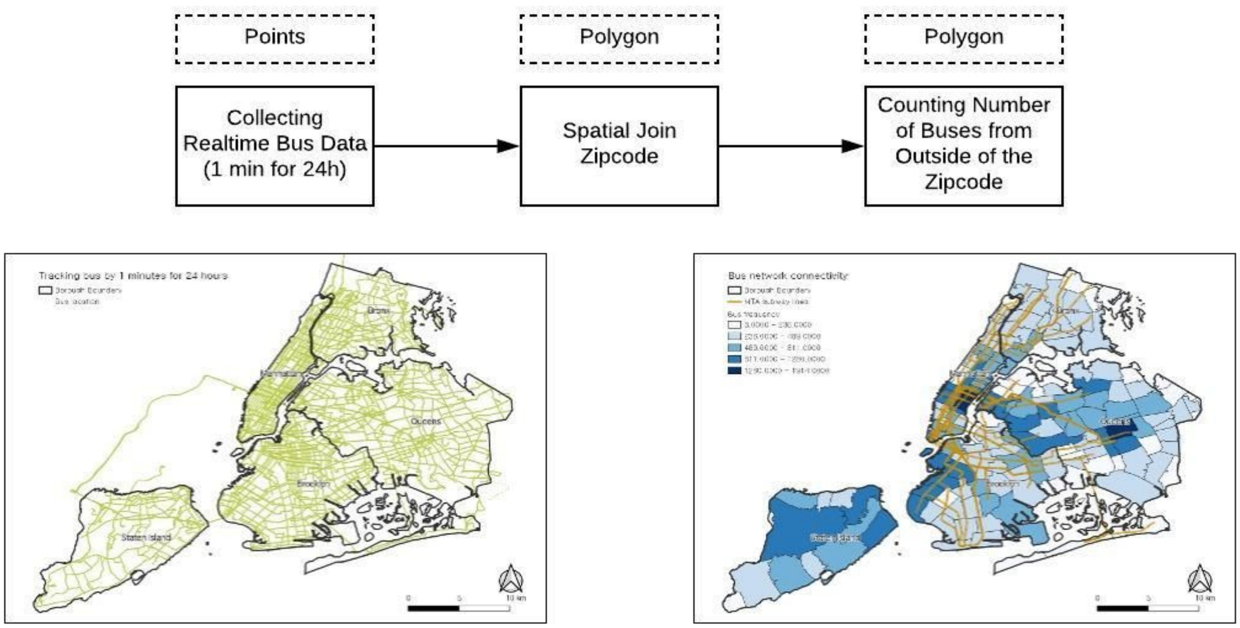
B. Existing bus supply analysis

The bus supply analysis is based on the MTA real-time bus data. To clarify the spatial distribution, we used zipcode as a spatial unit. Transit services have various schedule frequencies in terms of different subway lines and bus lines mainly due to various factors, including vehicle capacity, passenger demand, traffic, etc. The transit service frequency will be quantified through the computation of Z-score which allows us to see which services have fewer frequencies compared to the network average.

The method to collect bus data is described as follows:

- Collect real-time bus data from mta API about all buses operating at the time. The request is executing in every 1 minute for 24 hours.
- Spatial join with zipcode to represent the spatial distribution of the bus services.
- Count the number of buses by zipcode which is coming from outside of the zipcode.

According to the result, the map (Figure 2) shows Midtown Manhattan in Manhattan and Jamaica in Queens is the bus hub. Meanwhile, Bronx area, Astoria in Queens and outer area in Queens area has low bus frequency.



	name	bus_code	lat	lon	time
2110834	MTA NYCT_BX36	MTA NYCT_1000	40.841921	-73.878575	201904291533
2115114	MTA NYCT_BX36	MTA NYCT_1000	40.842496	-73.878232	201904291534
2119538	MTA NYCT_BX36	MTA NYCT_1000	40.842835	-73.878586	201904291535
2123876	MTA NYCT_BX36	MTA NYCT_1000	40.843832	-73.880337	201904291536
2129385	MTA NYCT_BX36	MTA NYCT_1000	40.844801	-73.882083	201904291537
2134924	MTA NYCT_BX36	MTA NYCT_1000	40.844879	-73.882240	201904291538
2138509	MTA NYCT_BX36	MTA NYCT_1000	40.845756	-73.884016	201904291539
2144264	MTA NYCT_BX36	MTA NYCT_1000	40.845756	-73.884016	201904291540

	name	bus_code	ZIPCODE	time
1311797	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291533
1311822	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291534
1311858	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291535
1311876	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291536
1311907	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291537
1311944	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291538
1311964	MTA NYCT_BX36	MTA NYCT_1000	10460	201904291539

Figure 2: Bus network frequency heatmap in NYC

Appendix 5

Synthetic Population Data Analysis for Specific Areas in NYC

The synthetic population trip data was provided by NYU C2SMART. The dataset is very large as it contains estimated personal trip data for 8+ million population in New York City. We initially selected 20 rows (origin, destination, population demographic characteristics and certain other relevant parameters) and then a smaller slice of dataset (10%) was used for the analysis.

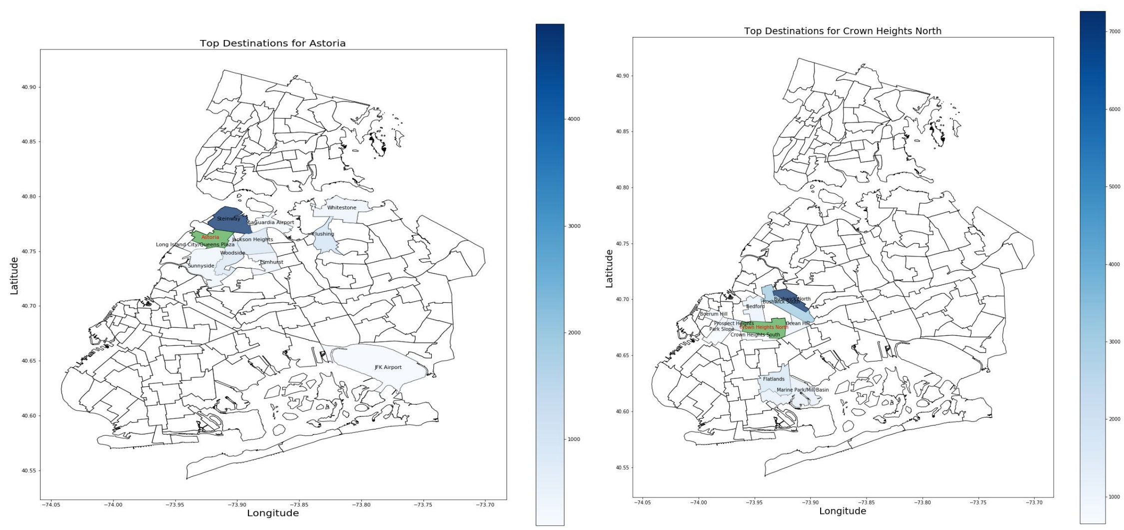


*Figure 1: Origin and destination for trips originating from Astori*a

The graph on the left shows the top destinations for trips originating from Astoria. The Astoria taxi zone is represented in green while top destinations include Steinway, Jackson Heights, and Flushing represented by the color scheme ranging from darker blue towards lighter blue. The graph on the right revealed that the top destinations for trips originating from Crown Heights North. The origin zone is colored in green while top destinations include Bushwick North, Bushwick South, and Flatlands are colored in blue. These suggest similar trend in comparing to the outcome of route selection.